SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853

TECHNICAL REPORT NO. 878

December 1989

ANTICIPATED BEHAVIOR OF PATH-FOLLOWING ALGORITHMS FOR LINEAR PROGRAMMING

by

S. Mizuno[1], M.J. Todd[2], and Y. Ye

[1]Research of this author was supported in part by a Grant-in-Aid for General Scientific Research of the Ministry of Education, Science and Culture, Japan.

[2]Research of this author was supported in part by NSF grant DMS-8904406 and ONR contract N00014-87-K0212.

Abstract

We provide a probabilistic analysis of the second-order term that arises in path-following algorithms for linear programming. We use this result to show that two specific such methods, a perfectly-centered algorithm of Sonnevend, Stoer and Zhao and a natural predictor-corrector algorithm, both require an "anticipated" number of iterations that is $O(n^{1/4}L)$.

1. Introduction

The aim of this paper is to try to explain the excellent practical behavior of interior-point algorithms for linear programming, as documented for instance in [1, 11, 12, 13, 17]. The best theoretical worst-case bound on the number of iterations is $O(\sqrt{n}\, L)$, for both path-following (e.g., Renegar [20], Gonzaga [5], Vaidya [24], Kojima, Mizuno and Yoshise [9], Monteiro and Adler [18, 19] and Mizuno [15]) and potential-reduction (e.g. Ye [26], Freund [4], and Kojima, Mizuno and Yoshise [10]) algorithms. Here n is the number of variables in a standard-form problem and L the size of the input, assumed integer. Alternatively, if a fairly central (see, e.g., Megiddo [14], Sonnevend [21], Bayer and Lagarias [2]) interior point feasible solution is known, L can be regarded as the number of additional bits of precision in the objective function value required. However, in practice the number of iterations required appears to grow much slower as a function of n; see in particular Lustig et al. [12], where a very good fit to a function linear in $\log n$ is obtained for a certain class of problems.

In this paper, we study the path-following algorithms in a probabilistic setting. We consider what we call the "anticipated" behavior of these methods. This notion was also considered in Gonzaga and Todd [7], where an "anticipated" $O(n^{1/4}L)$-iteration algorithm was described. The idea is that, at each iteration, we make an unrigorous but plausible assumption concerning the data of the problem, and then address the expected behavior, or, preferably, behavior which occurs with high probability (converging to 1 as $n \to \infty$), at that iteration. The anticipated number of iterations is then defined to be the number of iterations required if this high-probability behavior actually occurs at each iteration (or at least once every ten, say, iterations). A similar analysis was performed by Dantzig [3] for the simplex method. Note that we cannot conclude that the number of iterations is also small, in expectation or with high probability, when the method is applied to random problems drawn from a certain probability distribution, since the probabilistic assumptions made at each iteration may not be consistent with a single distribution on the original problem. Nevertheless, we feel such an analysis gives insight into the observed behavior of related algorithms.

With slight modifications, most path-following algorithms can be extended to solve positive semi-definite linear complementarity problems (LCP's) which include convex quadratic programming problems for instance; see, e.g., [8, 9, 19]. Here we confine our attention to methods for linear programming, because there does not appear to be a "natural" probabilistic assumption that will allow us to analyze the anticipated number of iterations in the case of the LCP.

Section 2 describes one iteration of a generic primal-dual path-following algorithm. Section 3 provides a probabilistic analysis of an important quantity that arises in an iteration of such a method. Section 4 applies this analysis to a conceptual algorithm where we assume that at the start of each iteration we have a perfectly centered primal-dual pair. This method was introduced and studied by Sonnevend, Stoer and Zhao [22], who show that, for certain rather restricted classes of problems, a worst-case complexity of $O(n^{\delta}L)$, where δ is strictly less than 1/2, can be obtained. Section 5 applies the result to a quite natural predictor-corrector method. In both cases, we show that the algorithms require $O(n^{1/4}L)$ anticipated number of iterations.

A companion paper [16] considers primal-dual interior-point algorithms that do not remain approximately centered as defined in the next section.

2. A generic path-following algorithm

In this section we justify our interest in the quantity studied in section 3 by describing a generic primal-dual path-following method. Consider the problem

$$(P) \qquad \begin{aligned} \min_{x} \quad & c^T x \\ Ax &= b \\ x &\geq 0 \end{aligned}$$

with dual

$$(D) \qquad \begin{aligned} \max_{y,s} \quad & b^T y \\ A^T y + s &= c \\ s &\geq 0, \end{aligned}$$

where A is $m \times n$, b and y are m-vectors, and c, s and x are n-vectors. The duality gap,

$c^Tx - b^Ty$, is easily seen to be equal to x^Ts for any feasible pair $(x,(y,s))$. Thus this pair is optimal if

$$
\begin{aligned}
Xs &= 0 \\
Ax &= b \\
A^Ty+s &= c \\
x \geq 0,\ \ & s \geq 0,
\end{aligned}
\tag{1}
$$

where X is the diagonal matrix with the components of x down its diagonal, $X = \mathrm{diag}(x)$. If both (P) and (D) have interior feasible solutions (with $x > 0$ and $s > 0$), then, for any $\mu \geq 0$, there will be a solution (unique if $\mu > 0$) to

$$
\begin{aligned}
Xs &= \mu e \\
Ax &= b \\
A^Ty+s &= c \\
x \geq 0,\ \ & s \geq 0,
\end{aligned}
\tag{2}
$$

If $\mu = 0$, such a solution satisfies (1) and hence is optimal; if $\mu > 0$, we must have $x > 0$ and $s > 0$ and x and (y,s) solve barrier problems naturally associated with (P) and (D); see, for example, Megiddo [14] and Bayer and Lagarias [2]. We call a solution to (2) (perfectly) centered. It is hard to find such a solution; hence one often works with (α-) approximately centered pairs, where the first equation is relaxed to

$$\|Xs - \mu e\| \leq \alpha\mu \tag{3}$$

for some $0 < \alpha < 1$. We use $\|\cdot\|$ for the Euclidean norm in this paper, while $\|\cdot\|_\infty$ is the ℓ_∞-norm. An approximateiy centered pair with $\mu > 0$ also has $x > 0$, $s > 0$. Note that the duality gap of such a pair is $x^Ts = e^TXs$, which lies in $\left[\left(1 - \frac{\alpha}{\sqrt{n}}\right)\mu n,\ \left(1 + \frac{\alpha}{\sqrt{n}}\right)\mu n\right]$. Hence it is natural, given an approximately centered pair (x, (y,s)) corresponding to $\mu > 0$, to seek a new approximately centered pair $(x^+,(y^+,s^+))$ corresponding to a smaller value $\mu^+ \in [0,\mu)$.

The Newton step to such a new centered pair is $d = (d_x,d_y,d_s)$, where

$$\begin{aligned} &Sd_x + Xd_s = \mu^+ e - Xs, \\ &Ad_x = 0, \ \ A^T d_y + d_s = 0; \end{aligned} \tag{4}$$

here $S = \text{diag}(s)$. Let $r = (XS)^{-1/2}(\mu^+ e - Xs)$, $p = (X^{-1}S)^{1/2}d_x$, and $q = (XS^{-1})^{1/2}d_s$. Then (4) can be rewritten:

$$\begin{aligned} &p + q = r \\ &p \in \text{null space}(A(XS^{-1})^{1/2}), \ \ q \in \text{row space}(A(XS^{-1})^{1/2}). \end{aligned} \tag{5}$$

Thus p is the projection of r into this null space, and q the projection of r into the complementary subspace.

Now let $x^+ = x + d_x$, $y^+ = y + d_y$ and $s^+ = s + d_s$. The following result is a slight extension of lemma 1 of [15]. We include a proof for completeness.

<u>Proposition</u> <u>1</u>. Let $X^+ = \text{diag}(x^+)$ and $P = \text{diag}(p)$, with the notation as above. Then

$$\begin{aligned} &X^+ s^+ - \mu^+ e = Pq, \ \text{and} \\ &(x^+)^T s^+ = n\mu^+ \end{aligned} \tag{6}$$

Moreover, if $\|Pq\| \le \alpha\mu^+$, then x^+ and (y^+, s^+) are α-approximately centered.

<u>Proof</u>. It is immediate that $Ax^+ = b$ and $A^T y^+ + s^+ = c$. If D_x denotes $\text{diag}(d_x)$, then from (4) we get

$$\begin{aligned}(X + \beta D_x)(s + \beta d_s) &= Xs + \beta S d_x + \beta X d_s + \beta^2 D_x d_s \\ &= (1-\beta)Xs + \beta\mu^+ e + \beta^2 D_x d_s \\ &= (1-\beta)Xs + \beta\mu^+ e + \beta^2 Pq.\end{aligned}$$

With $\beta = 1$, this gives $X^+s^+ - \mu^+e = Pq$ as desired, and then multiplying by e^T gives the second equation of (6) since $e^TPq = p^Tq = 0$.

Now if $||Pq|| \leq \alpha\mu^+$, then $\beta\mu^+e + \beta^2Pq \geq 0$ for all $\beta \in [0,1]$. Since x and (y,s) are α-approximately centered with $\mu > 0$, $Xs > 0$. Thus $(X + \beta D_x)(s + \beta d_s) \geq 0$ for all $\beta \in [0,1]$, with strict inequality if $\beta < 1$. But then $x^+, s^+ \geq 0$, for otherwise there would be some $\beta \in (0,1)$ with a component of $x + \beta d_x$ or $s + \beta d_s$ zero. Hence x^+ and (y^+, s^+) are α-approximately centered if $||Pq|| \leq \alpha\mu^+$.

Thus the crucial error, which we need to control if we want an approximately centered pair, is $||Pq||$; we need to "mind our p's and q's." Note that, from the proof above, Pq is the second-order term D_xd_s.

The sizes of p and q depend on that of $r = (XS)^{-1/2}(\mu e - Xs - (\mu-\mu^+)e)$; hence if $\mu-\mu^+$ is kept suitably small, we can guarantee that the new pair is approximately centered. By (6) however, the new duality gap is proportional to μ^+, which we therefore want to be as small as possible.

Hence studying the behavior of $||Pq||$ may allow us to decrease μ^+, and thus the duality gap, faster.

Many papers (e.g. [8, 9, 18, 19]) describe how to modify (P) and (D) so that an initial approximately centered pair with $x^Ts \leq 2^{O(L)}$ can easily be obtained. Moreover, if such a pair with $x^Ts \leq 2^{-2L}$ is known, exact solutions to (P) and (D) can be obtained in $O(n^3)$ arithmetic operations, as described in [9,19]. Hence if we can guarantee (or anticipate) that

$$\mu^+ \leq \left(1 - \frac{1}{f(n)}\right)\mu$$

for some function f, we will have an algorithm requiring $O(f(n)L)$ iterations in the worst-case (or anticipated case). Indeed, if we can guarantee (or anticipate) that this inequality holds at least once in every fixed number of iterations with $\mu^+ \le \mu$ at the other iterations, then the conclusion remains valid.

Finally we consider the asymptotic behavior of $\|Pq\|$ if x and (y,s) converge to nondegenerate optimal solutions. Then, if we suppose that the first m components of the optimal x are positive, the null space of $A(XS^{-1})^{1/2}$ converges to the space spanned by the last n-m unit vectors, and its row space to that spanned by the first m unit vectors. Hence $\|Pq\|/\|r\|^2$ converges to zero in this case.

3. Probabilistic Analysis of Pq

Let $\bar{A}$ be $m \times n$ and let $U \subseteq R^n$ be the null space of $\bar{A}$. For a given $r \in R^n$, let p denote the projection of r into U and q its projection into the orthogonal complement of U, so that $p+q = r$. Our interest is in the norm of Pq, where $P = \mathrm{diag}(p)$. As we have seen in section 2, this is the size of the second-order term in several path-following methods.

Mizuno [15] established (see also [9,19] for related results)

Proposition 2. With the notation above

$$\|Pq\| \le \frac{\sqrt{2}}{4} \|r\|^2. \tag{7}$$

This bound cannot be improved by much in the worst case, since it is possible to have

$r = e = (1,1,\dots,1)^T$,

$p = (\frac{1}{2}, \frac{1}{2},\dots,\frac{1}{2}, \alpha)^T$, and

$q = (\frac{1}{2}, \frac{1}{2},\dots,\frac{1}{2}, \beta)^T$, where

$\alpha = (1 + \sqrt{n})/2$ and $\beta = (1 - \sqrt{n})/2$, in which case $||Pq|| = \frac{1}{4}\sqrt{n(n-1)}$. Thus to obtain a tighter bound than (7) it is necessary to move to a probabilistic setting.

We will take r fixed, but assume that U is a random subspace of dimension $d := n-m$. This means that the distribution of U is invariant under orthogonal transformations. For instance, we can assume each entry of $\bar{A}$ is independently drawn from a standard normal distribution. See [23]. We assume throughout that $n \geq 4$. Our aim is to prove:

Theorem 1. Let $r \in R^n$ be fixed, and let U be a random d-subspace of R^n. Let p and q be the projections of r on U and $U^{\perp}$ respectively, and let P denote $\mathrm{diag}(p)$. Then if $\rho = ||r||_{\infty}/||r||$,

$$E(||Pq||) \leq \frac{||r||^2}{4}(\rho^2 + 3/n)^{1/2} \tag{8}$$

and for any $\epsilon > 0$,

$$\Pr\left\{||Pq|| \leq \frac{||r||^2}{4}(2\rho^2 + (6+\epsilon)/n)^{1/2}\right\} \to 1 \tag{9}$$

as $n \to \infty$.

We will generally use this result when $\rho = ||r||_{\infty}/||r|| = O(n^{-1/2})$; then it states that $||Pq||$ is $O(n^{-1/2})||r||^2$ in expectation and with high probability.

We divide the proof of the theorem into several lemmas. First we obtain the distribution of p. Because p and q are homogeneous of degree 1 in $||r||$, we assume henceforth without loss of generality that r is scaled so that

$$t := r/2 \text{ satisfies } ||t|| = 1, \ ||t||_{\infty} = \rho. \tag{10}$$

Let $[t,Y]$ be an orthogonal $n \times n$ matrix.

Lemma 1. We can write

$$p = (1 + \mu)t + \nu Yz, \tag{11}$$

where

$\frac{1+\mu}{2}$ has a beta distribution with parameters $\frac{d}{2}$ and $\frac{m}{2}$;

$\nu = \sqrt{1-\mu^2}$; and

z is uniformly distributed on the unit sphere in R^{n-1}.

Proof. Since p and q are orthogonal with $p+q = r$, p lies on the sphere of center $r/2 = t$ and radius $\|t\| = 1$. Thus p can be written in the form (11), with $\nu = \sqrt{1-\mu^2}$ and $\|z\| = 1$. We need to establish that μ and z have the given distributions.

Note that $\|p\|^2 = (1+\mu)^2 + \nu^2 = 2(1+\mu)$. However, we can obtain the distribution of $\|p\|^2$ directly. The invariance under orthogonal transformations implies that we can alternatively take U as a fixed d-subspace, say $\{x \in R^n: x_{d+1} = ... = x_n = 0\}$, and r uniformly distributed on a sphere of radius 2. Then r can be generated as

$$\left(\frac{2\lambda_1}{\|\lambda\|}, ..., \frac{2\lambda_n}{\|\lambda\|} \right)^T,$$

where $\lambda \sim N(0,I)$ in R^n (i.e., the components of λ are independent normal random variables with mean 0 and variance 1). But then

$$p = \left(\frac{2\lambda_1}{\|\lambda\|}, ..., \frac{2\lambda_d}{\|\lambda\|}, 0,...,0 \right)^T,$$

and $\|p\|^2 = 4(\lambda_1^2 + ... + \lambda_d^2)/(\lambda_1^2 + ... + \lambda_n^2)$. This has the distribution of four times a beta random variable with parameters $\frac{d}{2}$ and $\frac{m}{2}$ (see, e.g., Wilks [25]), which confirms the distribution of μ.

Now let W be an orthogonal matrix with $Wt = t$. W can be viewed as rotating the sphere with center t around its diameter from 0 to $2t = r$. The fact that p is the projection of r onto

$U = \{x\colon \bar{A}x = 0\}$ is equivalent to $\bar{A}p = 0$, $r - p = \bar{A}^T v$ for some v. But then $(\bar{A}W^T)Wp = 0$ and $r - Wp = Wr - Wp = (\bar{A}W^T)^T v$, so that Wp is the projection of r onto $U' = \{x\colon (\bar{A}W^T)x = 0\}$. If $\bar{A}$ has independent standard normal entries, so does $\bar{A}W^T$, so U' is also a random d-subspace. Thus Wp has the same distribution as p. But writing W as $YVY^T + tt^T$, where V is an arbitrary orthogonal matrix of order $n-1$, we see that z has the same distribtion as Vz. Since $||z|| = 1$, z is uniformly distributed on the unit sphere in R^{n-1}. □

From (10) we have

$$q = (1-\mu)t - \nu Yz, \text{ so that}$$
$$Pq = \nu^2 t^2 - 2\mu\nu TYz - \nu^2 (Yz)^2. \tag{12}$$

where $T := \mathrm{diag}(t)$ and t^2, $(Yz)^2$ denote the vectors whose components are the squares of those of t, Yz respectively.

Suppose for the moment that t is the first unit vector. Then we can take for Y the last $n-1$ columns of the identity matrix, so that TYz is zero and $(Yz)^2$ is zero in its first component. Hence $||Pq|| \geq \nu^2$ in this case. But from the distribution of μ it is easily shown (cf. [24]) that, if $d = m = n/2$, ν^2 has expectation $n/(n+2)$. Thus we cannot hope to have $E(||Pq||) = O(n^{-1/2})||r||^2$ for general r.

Returning to the general case, we find

$$||Pq||^2 = \nu^4 \sum t_j^4 + \nu^4 \sum ((Yz)_j)^4 + 4\mu^2\nu^2 ||TYz||^2 - 2\nu^4 ||TYz||^2 \tag{13}$$
$$- 4\mu\nu^3 \sum t_j^3 (Yz)_j + 4\mu\nu^3 \sum t_j ((Yz)_j)^3.$$

The first term is at most $\nu^4 \rho^2 \sum t_j^2 = \nu^4 \rho^2$, by recalling that $\max|t_j| = ||t||_\infty = \rho$. The third and fourth terms combine to give

$$2\nu^2(3\mu^2 - 1)||TYz||^2 \le 2\nu^2 \max\{0, 3\mu^2 - 1\}\rho^2,$$

since $||TYz||^2 \le ||t||_\infty^2 \, ||Yz||^2 = ||t||_\infty^2 = \rho^2$. Hence, if $\mu^2 \le \frac{1}{3}$, the first, third and fourth terms are at most ρ^2. If $\mu^2 > \frac{1}{3}$, they are at most $\rho^2\nu^2(\nu^2 + 2(3\mu^2 - 1) = \rho^2\nu^2(4-5\nu^2) = \rho^2(\frac{4}{5} - 5(\nu^2 - \frac{2}{5})^2) < \rho^2$. Hence, in either case,

$$\text{the first, third and fourth terms in (13) are at most } \rho^2. \tag{14}$$

Next, note that the fifth and sixth terms are odd functions of z. Since z has a symmetric distribution about 0 by Lemma 1,

$$\text{the fifth and sixth terms in (13) have expectation 0.} \tag{15}$$

We therefore have to consider $\sum((Yz)_j)^4$.

Lemma 2. Let

$$\gamma := \sum((Yz)_j)^4. \tag{16}$$

Then

$$\frac{3(n-2)}{n^2-1} \le E\gamma \le \frac{3}{n}\,. \tag{17}$$

Proof. Let y_j^T denote the jth row of Y. Since $[t,Y]$ is orthogonal, $||y_j|| = (1-t_j^2)^{1/2} =: \tau_j$. Then

$$E\gamma = \sum E(y_j^T z)^4 = \sum \tau_j^4 E(z^T y_j/||y_j||)^4.$$

Now $z^T y_j/||y_j||$ is distributed just like z_1 (by orthogonal invariance), and z_1 can be generated as $\lambda_1/||\lambda||$, where $\lambda \sim N(0,I)$ in R^{n-1}. Then z_1^2 has a beta distribution with $\frac{1}{2}$ and $\frac{n-2}{2}$ degrees of freedom, and

$$Ez_1^4 = \Gamma\left(\tfrac{1}{2} + 2\right)\Gamma\left(\tfrac{n-1}{2}\right)/\left[\Gamma\left(\tfrac{1}{2}\right)\Gamma\left(\tfrac{n-1}{2} + 2\right)\right]$$

$$= \frac{\frac{3}{2} \cdot \frac{1}{2}}{\frac{n+1}{2}\,\frac{n-1}{2}} = \frac{3}{n^2-1},$$

where Γ is the gamma function. See Wilks [25]. Thus

$$E\gamma = \frac{3}{n^2-1} \sum \tau_j^4$$

$$= \frac{3}{n^2-1} \sum(1 - t_j^2)^2$$

$$= \frac{3}{n^2-1} (n - 2 \sum t_j^2 + \sum t_j^4)$$

$$= \frac{3}{n^2-1} (n - 2 + \sum t_j^4).$$

Since $0 \leq \sum t_j^4 \leq 1$, (ii) follows. □

From (13)-(17), we have

$$E(\|Pq\|)^2 \leq \rho^2 + \tfrac{3}{n}, \quad \text{so}$$

$$E(\|Pq\|) \leq \left(\rho^2 + \tfrac{3}{n}\right)^{1/2},$$

and (8) follows. (Recall that we are assuming $\|r\| = 2$ without loss of generality.)

In order to prove (9), we must make a stronger statement about the odd-degree terms in (13) and show that γ in (16) concentrates around its mean. Suppose we can show that

$$\Pr\left\{\gamma \leq \frac{3+\eta}{n}\right\} \to 1 \tag{18}$$

as $n \to \infty$ for any $\eta > 0$. Let F be the event that $\gamma > (3+\eta)/n$, so that $\Pr(F) \to 0$. Now if $(\mu,z) \notin F$, then the first through fourth terms of (13) are at most $\rho^2 + (3+\eta)/n$. If for any such (μ,z), (13) were greater than $2\rho^2 + (6+2\eta)/n$, then for $(\mu, -z)$, (13) would be negative, a contradiction. (The even-degree terms stay the same while the odd-degree terms switch sign.) If we therefore choose $\eta = \epsilon/2$, we will have shown (9). It remains to prove (18).

In order to use Chebychev's inequality (see, e.g., [25]), we need to find the variance of γ. Lemma 2 gives tight bounds on $E\gamma$, so we concentrate on $E\gamma^2$. Now

$$\begin{aligned} E\gamma^2 &= E\Big(\sum (y_i^T z)^4\Big)\Big(\sum (y_j^T z)^4\Big) \\ &= \sum E(y_j^T z)^8 + 2\sum_{i<j} E\Big[(y_i^T z)^4 (y_j^T z)^4\Big] \qquad (19) \\ &= \Big(\sum \tau_j^8\Big) E(z^T y_1/\|y_1\|)^8 + 2\sum_{i<j} E\Big[(y_i^T z)^4 (y_j^T z)^4\Big], \end{aligned}$$

where $\tau_j^2 = 1 - t_j^2 = \|y_j\|^2$. The first expectation is that of β^4, where β has the beta distribution with $\frac{1}{2}$ and $\frac{n-2}{2}$ degrees of freedom. Now

$$E\beta^4 = \frac{\frac{7}{2}\cdot\frac{5}{2}\cdot\frac{3}{2}\cdot\frac{1}{2}}{\frac{n+5}{2}\cdot\frac{n+3}{2}\cdot\frac{n+1}{2}\cdot\frac{n-1}{2}} \le \frac{105}{n^4}$$

([24]). Hence the first term is at most $105/n^3$.

To estimate the second expectation, which we denote σ_{ij}, we perform a rotation on z so that the ith and jth rows of $[t,Y]$ are

$$\begin{bmatrix} t_i & \tau_i & 0 & 0 & \cdots & 0 \\ t_j & \nu_j & \theta_j & 0 & \cdots & 0 \end{bmatrix}.$$

If $\tau_i = 0$, then $y_i = 0$ and $\sigma_{ij} = 0$. Otherwise, orthogonality gives $\nu_j = -t_i t_j/\tau_i$ and we find

$$
\begin{aligned}
\sigma_{ij} &= E[(\tau_i z_1)^4 (\nu_j z_1 + \theta_j z_2)^4] \\
&= E[(t_i t_j)^4 z_1^8] + 6E[(t_i t_j)^2 \tau_i^2 \theta_j^2 z_1^6 z_2^2] + E[\tau_i^4 \theta_j^4 z_1^4 z_2^4] \\
&\le (t_i t_j)^4 E z_1^8 + 6(t_i t_j)^2 E(z_1^6 z_2^2) + E(z_1^4 z_2^4),
\end{aligned} \tag{20}
$$

where the omitted terms are of odd degree in z_1 and z_2 and hence have expectation zero. As above $Ez_1^8 \le 105/n^4$. To control the first two terms in (20), we use

Lemma 3. $\sum_{i<j} (t_i t_j)^4 \le \sum_{i<j} (t_i t_j)^2 \le \frac{1}{2}$.

Proof. The first inequality is trivial as $(t_i t_j)^2 \le 1$. Now the maximum of $\sum_{i<j} (t_i t_j)^2$ over $\sum t_j^2 = 1$ is attained at some $\bar{t}$, and by Lagrange's theorem, $\bar{t}_j \sum_{k \ne j} \bar{t}_k^2 = \lambda \bar{t}_j$ for some λ and for all j. It follows that $\bar{t}_j^2 = 0$ or $1 - \lambda$ for all j, so that k of the $\bar{t}_j^2$ are $\frac{1}{k}$ and the rest zero, whence $\sum_{i<j} (t_i t_j)^2 = \frac{k(k-1)}{2} \cdot \frac{1}{k^2} \le \frac{1}{2}$. □

The proof of the final lemma we need is technical and deferred to the appendix.

Lemma 4. $E(z_1^6 z_2^2) \le 15/n^4$ and $E(z_1^4 z_2^4) = 9/[(n^2-1)(n+3)(n+5)]$.

Now putting together (19), (20) and these two lemmas, we arrive at

$$
E\gamma^2 \le \frac{105}{n^3} + 2 \cdot \frac{1}{2} \cdot \frac{105}{n^4} + 2 \cdot 6 \cdot \frac{1}{2} \cdot \frac{15}{n^4} + \frac{n(n-1) \cdot 9}{(n^2-1)(n+3)(n+5)}
$$

so that, using (17), we find

$$
\mathrm{Var}(\gamma) = E\gamma^2 - (E\gamma)^2 = O\left(\frac{1}{n^3}\right)
$$

Finally, Chebychev's inequality using our estimates of the expectation and variance of γ establishes (18), and this at last completes the proof of the theorem.

4. A perfectly-centered path-following algorithm

Here we apply the results of the previous section to a method that is equivalent to Algorithm 1 of Sonnevend, Stoer, and Zhao [22]. In a sense, this is a conceptual algorithm, since it assumes that each iteration begins and ends with an exactly centered pair. However, Sonnevend, Stoer, and Zhao argue that if an α-approximately centered pair is available for a suitably small α, then a centered pair (to within working precision) can be found in a fixed number of iterations of Newton's method, depending only on α and the precision required. This follows from the quadratic convergence results (with explicit constants) that have been obtained -- see, e.g., Gonzaga [6].

Let us assume therefore that a perfectly centered pair (x,(y,s)) for (P) and (D), satisfying (2), is available. We then solve

$$Sd_x + Xd_s = -\theta\mu e$$

$$Ad_x = 0, \ \ A^Td_y + d_s = 0$$

for d. Note that (4) gives collinear directions $d(\mu^+)$ for any μ^+ when $Xs = \mu e$ -- we have chosen

$$\mu^+ = (1\text{-}\theta)\mu. \tag{21}$$

We choose θ as large as possible (μ^+ as small as possible) while ensuring that the new iterates are α-approximately centered for some fixed suitably small α. Then a small number of recentering steps will yield (to working precision) a perfectly centered pair corresponding to μ^+.

From Proposition 1, in order that the new iterates are α-approximately centered, we need

$$||Pq|| \leq \alpha\mu^+ = \alpha(1\text{-}\theta)\mu. \tag{22}$$

Here p and q are the projections onto the null space of $A(XS^{-1})^{1/2}$ and onto its orthogonal complement of

$$r = (XS)^{-1/2}(\mu^+ e - Xs)$$

$$= -\theta\mu^{1/2}e.$$

Let $\beta = ||Pq||/||r||^2 = ||Pq||/n\theta^2\mu$; then (22) holds iff

$$n\beta\theta^2 + \alpha\theta - \alpha \leq 0,$$

and we can therefore choose

$$\theta = \left(\sqrt{\alpha^2 + 4n\alpha\beta} - \alpha\right)/2n\beta$$

$$= 2\alpha/\left(\sqrt{\alpha^2 + 4n\alpha\beta} + \alpha\right).$$

From Proposition 2, β is at most a constant in the worst case, so that

$$\theta = \Omega(n^{-1/2}). \tag{23}$$

However, since $||r||_\infty/||r|| = 1/\sqrt{n}$, if the probabilistic assumptions of section 3 hold, then with probability approaching 1 as $n \to \infty$, $\beta \leq \frac{3}{4}n^{-1/2}$ by Theorem 1, and we have

$$\theta = \Omega(n^{-1/4}). \tag{24}$$

From the remarks at the end of section 2, (21), (23) and (24) yield

<u>Theorem</u> <u>2</u>. The algorithm of this section requires an anticipated number of iterations that is $O(n^{1/4}L)$, with a worst-case bound of $O(n^{1/2}L)$. □

5. A predictor-corrector algorithm

In this section we analyze an algorithm that takes a single "corrector" step to the central path after each "predictor" step to decrease μ. Let $N(\alpha)$ denote the set of pairs that are α-approximately centered for some μ. We work with (nearly centered) pairs in $N(\frac{1}{2})$ and (very nearly centered) pairs in $N(\frac{1}{4})$.

Given a (very nearly centered) pair $(x,(y,s)) \in N(\frac{1}{4})$, an iteration proceeds as follows:

Predictor step:

Let $d = (d_x, d_y, d_s)$ solve

$$\begin{aligned} Sd_x + Xd_s &= -Xs \\ Ad_x = 0, \ \ A^T d_y + d_s &= 0 \end{aligned} \tag{25}$$

and let

$$(x', y', s') = (x + \theta d_x, y + \theta d_y, s + \theta d_s) \tag{26}$$

where

$$\theta = \max\{\gamma \in [0,1]: \ (x + \gamma d_x, \ (y + \gamma d_y, \ s + \gamma d_s)) \in N(\tfrac{1}{2})\}. \tag{27}$$

Corrector step:

Let $d' = (d'_x, d'_y, d'_s)$ solve

$$S'd'_x + X'd'_s = \mu^+ e - X's' \tag{28}$$

$$Ad'_x = 0, \ \ A^T d'_y + d'_s = 0,$$

where

$$\mu^+ = (x')^T s'/n, \tag{29}$$

and set

$$(x^+, y^+, s^+) = (x' + d'_x, \ y' + d'_y, \ s' + d'_s). \tag{30}$$

Lemma 5. The pair $(x^+, (y^+, s^+))$ lies in $N(\frac{1}{4})$.

Proof. By the choice of θ in (27), we have $(x', (y', s')) \in N(\frac{1}{2})$. Then applying Proposition 1 to the corrector step, we have

$$\|X^+s^+ - \mu^+e\| = \|P'q'\|,$$

where p' and q' correspond to the data x', y' and s'. Now from Proposition 2,

$$\begin{aligned}\|P'q'\| &\le \frac{\sqrt{2}}{4}\|r'\|^2 \\ &= \frac{\sqrt{2}}{4}\|(X'S')^{-1/2}(\mu^+e - X's')\|^2 \\ &\le \frac{\sqrt{2}}{4}\left(\frac{1}{\sqrt{\mu^+(1-\frac{1}{2})}}(1/2\ \mu^+)\right)^2 \\ &< \frac{\mu^+}{4},\end{aligned}$$

which gives the desired result. □

Note that $(s')^T d'_x + (x')^T d'_s = 0$ from (28), so that (30) implies

$$(x^+)^T s^+ = n\mu^+ = (x')^T s'.$$

Similarly, using (25) and (26) gives

$$\begin{aligned}n\mu^+ &= (x')^T s' = x^T s - \theta x^T s \\ &= (1-\theta)n\mu, \text{ or}\end{aligned}$$

$$\mu^+ = (1 - \theta)\mu, \tag{31}$$

so that we again want to find lower bounds on θ.

Lemma 6. $\theta \geq \bar{\gamma} := \min\left\{\frac{1}{2}, \left(\frac{\mu}{8||Pq||}\right)^{1/2}\right\}$, where p and q correspond to the predictor step.

Proof. For any $\gamma \geq 0$,

$$\begin{aligned}||(X + \gamma D_X)(s + \gamma d_S) - (1 - \gamma)\mu e|| &= ||Xs + \gamma(Sd_X + Xd_S) + \gamma^2 D_X d_S - (1 - \gamma)\mu e|| \\ &= ||(1 - \gamma)(Xs - \mu e) + \gamma^2 D_X d_S|| \\ &\leq \frac{1}{4}(1 - \gamma)\mu + \gamma^2 ||Pq||.\end{aligned}$$

For $\gamma = \bar{\gamma}$, this quantity is at most

$$\begin{aligned}\frac{1}{4}(1 - \gamma)\mu + \mu/8 &\leq \frac{1}{4}(1 - \gamma)\mu(1 + 1/2(1 - \gamma)) \\ &\leq \frac{1}{2}(1 - \gamma)\mu.\end{aligned}$$

Since it is easy to see that $(x + \gamma d_X)^T(s + \gamma d_S) = (1 - \gamma)n\mu$, this shows that $\bar{\gamma}$ gives a pair in $N(\frac{1}{2})$, so that $\theta \geq \bar{\gamma}$. □

Now $r = (Xs)^{-1/2}(-Xs)$ in the predictor step, from which $||r||^2 = n\mu$ follows. Thus Proposition 2 yields $(\mu/8||Pq||)^{1/2} \geq 8^{-1/4}n^{-1/2}$, and Lemma 6 then implies that

$$\theta = \Omega(n^{-1/2}) \tag{32}$$

in the worst-case. However, we can also deduce that $||r||_\infty = ||Xs||_\infty^{1/2} \leq (3\mu/2)^{1/2} = O(n^{-1/2})||r||$. Hence, if the probabilistic assumptions of section 3 hold,

$$||Pq||/\mu = O(n^{1/2})$$

and

$$\theta = \Omega(n^{-1/4}) \tag{33}$$

with probability converging to 1 as $n \rightarrow \infty$.

As in section 4, we thus obtain from (31)-(33)

Theorem 3. The predictor-corrector algorithm requires $O(n^{1/2}L)$ iterations in the worst case, but the anticipated number of iterations is $O(n^{1/4}L)$.

References

1. I. Adler, N. Karmarkar, M.G.C. Resende, and G. Veiga, "An implementation of Karmarkar's algorithm for linear programming," Mathematical Programming 44 (1989).

2. D. Bayer and J.C. Lagarias, "The nonlinear geometry of linear programming: I. Affine and projective scaling trajectories" and "II. Legendre transform coordinates and central trajectories," Transactions of the American Mathematical Society 314 (1989), 499-526 and 527-581.

3. G.B. Dantzig, "Expected number of steps of the simplex method for a linear program with a convexity constraint," Technical Report SOL 80-3, Department of Operations Research, Stanford University, Stanford, CA, 1980.

4. R. Freund, "Polynomial-time algorithms for linear programming based only on primal scaling and projected gradients of a potential function," manuscript, Sloan School of Management, M.I.T., Cambridge, MA, 1988.

5. C. C. Gonzaga, "An algorithm for solving linear programming in $O(n^3L)$ operations," in: Progress in Mathematical Programming (N. Megiddo, ed.), Springer-Verlag, Berlin, 1-28, 1988.

6. C.C. Gonzaga, "Large-steps path-following methods for linear programming: barrier function method," Technical Report, Department of Systems Engineering and Computer Sciences, COPPE-Federal University of Rio de Janeiro, Brazil, July 1989.

7. C. C. Gonzaga and M.J. Todd, "An $O(\sqrt{n}\, L)$-iteration large-step primal-dual affine algorithm for linear programming," Technical Report No. 862, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York, 1989.

8. M. Kojima, S. Mizuno and A. Yoshise, "A primal-dual interior point method for linear programming," in: Progress in Mathematical Programming (N. Megiddo, ed.), Springer Verlag, Berlin, 29-47, 1988.

9. M. Kojima, S. Mizuno and A. Yoshise, "A polynomial algorithm for a class of linear complementarity problems," Mathematical Programming 44 (1989), 1-26.

10. M. Kojima, S. Mizuno and A. Yoshise, "An $O(\sqrt{n}\, L)$ iteration potential reduction algorithm for linear complementarity problems," Research Report 8-217, Department of Information Sciences, Tokyo Institute of Technology, Tokyo, Japan, 1988.

11. I.J. Lustig, R.E. Marston and D.F. Shanno, "Computational experience with a primal-dual interior point method for linear programming," Technical Report SOR 89-17, Department of Civil Engineering and Operations Research, Princeton University, Princeton, NJ, 1989.

12. I.J. Lustig, D.F. Shanno and J.W. Gregory, "The primal-dual interior point method on a Cray supercomputer," Technical Report SOR 89-21, Department of Civil Engineering and Operations Research, Princeton University, Princeton, NJ, 1989.

13. K.A. McShane, C.L. Monma and D.F. Shanno, "An implementation of a primal-dual interior point method for linear programming," ORSA Journal on Computing 1 (1989), 70-83.

14. N. Megiddo, "Pathways to the optimal set in linear programming," in: Progress in Mathematical Programming (N. Megiddo, ed.), Springer-Verlag, Berlin, 131-158, 1988.

15. S. Mizuno, "A new polynomial time method for a linear complementarity problem," Technical Report 16, Department of Industrial Engineering and Management, Tokyo Institute of Technology, Tokyo, Japan, 1989.

16. S. Mizuno, M.J. Todd and Y. Ye, "Anticipated behavior of long-step algorithms for linear programming," in preparation.

17. C.L. Monma and A.J. Morton, "Computational experience with a dual affine variant of Karmarkar's method for linear programming," Operations Research Letters 6 (1987), 261-167.

18. R.C. Monteiro and I. Adler, "Interior path following primal-dual algorithms, Part I: Linear programming," Mathematical Programming 44 (1989), 27-41.

19. R.C. Monterio and I. Adler, "Interior path following primal-dual algorithms, Part II: Convex quadratic programming," Mathematical Programming 44 (1989), 43-66.

20. J. Renegar, "A polynomial-time algorithm based on Newton's method for linear programming," Mathematical Programming 40 (1988), 59-93.

21. G. Sonnevend, "An analytical centre for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming," Lecture Notes in Control and Information Sciences 84, Springer-Verlag, New York, 866-876, 1986.

22. G. Sonnevend, J. Stoer and G. Zhao, "On the complexity of following the central path by linear extrapolation in linear programs," to appear in the Proceedings of the 13th Symposium on Operations Research (1989 Ulm), ed. U. Rieder, 1989.

23. M.J. Todd, "Probabilistic models for linear programming," Technical Report 836, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY, 1989.

24. P.M. Vaidya, "An algorithm for linear programming which requires $O(((m+n)n^2 + (m+n)^{1.5}n)L)$ arithmetic operations," preprint, AT&T Bell Laboratories, Murray Hill, New Jersey, 1987, to appear in Mathematical Programming.

25. S.S. Wilks, Mathematical Statistics, John Wiley, New York, 1962.

26. Y. Ye, "An $O(n^3L)$ potential reduction algorithm for linear programming," manuscript, Department of Management Sciences, The University of Iowa, Iowa City, Iowa (1988, revised 1989), to appear in Mathematical Programming.

Appendix

Here we prove

Lemma 4. If $z \in \mathbb{R}^{n-1}$ is uniformly distributed on the unit sphere, then

$$E(z_1^6 z_2^2) \le \frac{15}{n^4} \quad \text{and} \quad E(z_1^4 z_2^4) = \frac{9}{(n^2-1)(n+3)(n+5)} .$$

Proof. We can write $z = \lambda/\|\lambda\|$ where $\lambda \sim N(0,I)$. Then $z_i^2 = \lambda_i^2/(\lambda_1^2 + \lambda_2^2 + \lambda_3^2 + \dots + \lambda_{n-1}^2)$, $i = 1,2$. Let $\lambda_1^2 =: x_1$, $\lambda_2^2 =: x_2$, and $\lambda_3^2 + \dots + \lambda_{n-1}^2 =: x_3$, where x_1, x_2 and x_3 are gamma random variables with degrees of freedom $\frac{1}{2}, \frac{1}{2}$ and $(n-3)/2$ respectively. Thus

$$E(z_1^4 z_2^4) = \int_0^\infty \int_0^\infty \int_0^\infty \frac{1}{(\Gamma(\frac{1}{2}))^2 \Gamma(\frac{n-3}{2})} \cdot x_1^{-1/2} \cdot x_2^{-1/2} \cdot x_3^{\frac{n-5}{2}} \cdot e^{-x_1-x_2-x_3} \cdot \frac{x_1^2 x_2^2}{(x_1+x_2+x_3)^4} \, dx_1 dx_2 dx_3 .$$

We now use the change of variables

$$\begin{aligned} x_1 &= \sin^2\theta \sin^2\phi \, y & \qquad 0 &\le \theta \le \pi/2 \\ x_2 &= \cos^2\theta \sin^2\phi \, y & \qquad 0 &\le \phi \le \pi/2 \\ x_3 &= \cos^2\phi \, y & \qquad 0 &\le y. \end{aligned}$$

The Jacobian of the transformation is

$$4 \sin\theta \cos\theta \sin^3\phi \cos\phi \, y^2,$$

so recalling that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, we find

$$E(z_1^4 z_2^4) = \frac{1}{\pi\Gamma(\frac{n-3}{2})} \int_0^{\pi/2} \int_0^{\pi/2} \int_0^{\infty} 4 \sin^4\theta \cos^4\theta \sin^9\phi \cos^{n-4}\phi \, y^{\frac{n-3}{2}} e^{-y} dy \, d\theta \, d\phi$$

$$= \frac{4}{\pi\Gamma(\frac{n-3}{2})} \left(\int_0^{\infty} y^{\frac{n-3}{2}} e^{-y} dy \right) \left(\int_0^{\pi/2} \sin^4\theta \cos^4\theta \, d\theta \right) \left(\int_0^{\pi/2} \sin^9\phi \cos^{n-4}\phi \, d\phi \right)$$

$$= \frac{4}{\pi\Gamma(\frac{n-3}{2})} \cdot \Gamma\left(\frac{n-1}{2}\right) \cdot \frac{3}{256} \pi \cdot \frac{384}{(n-3)(n-1)(n+1)(n+3)(n+5)}$$

$$= \frac{9}{(n^2-1)(n+3)(n+5)} \text{, as desired.}$$

Using exactly the same transformations,

$$E(z_1^6 z_2^2) = \frac{4}{\pi\Gamma(\frac{n-3}{2})} \left(\int_0^{\infty} y^{\frac{n-3}{2}} e^{-y} dy \right) \left(\int_0^{\pi/2} \sin^6\theta \cos^2\theta \, d\theta \right) \left(\int_0^{\pi/2} \sin^9\phi \cos^{n-4}\phi \, d\phi \right)$$

$$= \frac{15}{(n^2-1)(n+3)(n+5)} . \quad \square$$